Journal of Computer Science

Research Article

Early Plant Disease Detection Using Graph Isomorphic Networks: Enhancing Crop Yield Through Leaf Analysis

[1]D. Sumathi, [2]Sreejyothsna Ankam, [3]Pravin Prakash Adivarekar, [4]Kasturi Sai Sandeep, [5]Gomathi R., [6]R. Shobarani, [7]S. Karpaga Iswarya and [8]V. Bhoopathy

[1]*Department of Computer Science and Engineering, Alliance University, Bangalore, India*
[2]*Department of Computer Science and Engineering, GMR Institute of Technology, Rajam, India*
[3]*Department of Computer Science and Engineering, A. P. Shah Institute of Technology, Thane, India*
[4]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India*
[5]*Department of Artificial Intelligence and Data Science, Bannari Amman Institute of Technology, Sathyamangalam, India*
[6]*Department of Computer Science and Engineering, Dr. M. G. R. Educational and Research Institute, Chennai, India*
[7]*Department of Computer Science and Engineering, Nehru Institute of Technology, Coimbatore, India*
[8]*Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India*

Article history
Received: 04-12-2024
Revised: 18-01-2025
Accepted: 22-01-2025

Corresponding Author:
V. Bhoopathy
Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India
Email: v.bhoopathy@gmail.com

Abstract: The economy of Tanzania is mostly driven by agriculture. Disease is one of the reasons that contributes to the low production of staple foods like cassava and maize, alongside climate change. Loss of income and food security are the results. In order to detect the diseases early, preventative measures are required. A potential option for farmers could be the use of image processing tools to identify plant diseases on leaves. Implementing the existing method of disease detection, which involves an expert using their naked eyes, on a large farm is a laborious and time-consuming process. This study provides a comprehensive overview of recent research in image processing by reviewing methods for identifying plant diseases in their leaves or fruits and the corresponding machine learning models for disease classification. This study examines issues in the identification of plant diseases, pertinent to agriculture-dependent nations like Tanzania and India. Presenting the present state of the art, elucidating the steps done during the image processing stage, and assessing the pros and cons of each technique as well as the effectiveness of the machine learning model used for disease classification are the primary goals of the work. Among the preprocessing and resampling techniques, the evaluation's results show that GIN-based approach for resampling, in conjunction with contrast limited adaptive histogram equalization (CLAHE), achieved the best results, with an average F1-score of 95.65% and a classification accuracy of 95.62%. The study concludes with a generic process for a disease detection system, which may be broken down into individual components as needed.

Keywords: Graph Isomorphic Network (GIN), Graph Neural Network (GNN), Plant Disease Detection, PlantDoc Dataset, Image Processing, Adaptive Histogram Equalization (AHE)

Introduction

Agriculture has played a significant role in India's economic progress. If agricultural damage significantly reduced productivity, the economy would suffer. Leaves are the first to display symptoms of illness due to their fragility (Sharma *et al.*, 2023). It is important to keep an eye on crops for signs of disease from the time they are seedlings all the way to harvest. Historically, plant disease monitoring relied on the time-consuming and error-prone practice of naked-eye inspection, which required experts to physically oversee crop fields. In recent years, numerous approaches have been employed to develop automated and semi-automatic systems that can identify plant diseases. Compared to the traditional method of farmers' manual observation, these methods are faster, cheaper, and more accurate so far. As a result, there is a pressing need to provide technical solutions that can identify plant diseases more independently (Khalid & Karan, 2023). Whether on enormous commercial farms or small subsistence farms, crop production is essential to human survival. Pathogens such as bacteria, fungi, viruses, and others have persisted throughout the history of this vital sector. By persistently endangering the very essence of agriculture, these invisible enemies erode food security and sustainability. On a worldwide basis, plants are largely responsible for providing food (Sunil *et al.*, 2022). However, they are susceptible to infections due to a variety of environmental factors, which greatly reduces their productivity. The increase in plant diseases has a detrimental effect on agricultural production. Failure to promptly detect plant diseases will exacerbate food shortages. Crop failure is directly caused by plant-eating pests, weeds, and diseases, which in turn lead to economic and production losses.

The host plant, an ideal surrounding environment, and the infectious agent all have a role in the development of plant diseases. The plant disease triangle seen in Figure (1) is a result of these variables. When a plant gets sick, the symptoms usually start at the base and work their way up. After infecting a crop, many plant diseases spread to other parts of the crop. Thus, it is crucial to keep an eye on crops on a frequent basis, as early disease treatment can help stop their spread.

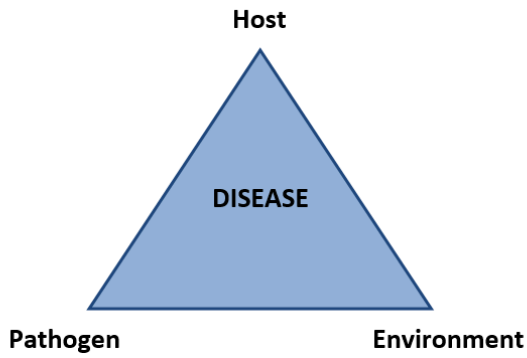

Fig. 1: Proposed model architecture

Improving agricultural yields and quality is difficult due to a multitude of other factors, such as the increase in greenhouse gas emissions and the use of chemical fertilizers in modern farming methods. Infected plants often show apparent symptoms, such as lesions, on their leaves, trunks, flowers, and fruits. It is common practice to use a single visual template for all insect or disease habitats when evaluating abnormalities (Sharma, 2024). It is common for plant diseases to spread through the leaves of infected plants, and it is often the leaves that show the earliest signs of a disease's prophetic significance. Experts in agricultural and plant pathology sometimes make house calls to farmers to provide connection-based diagnoses of pests and diseases affecting their crops. All at once, this approach is lofty, useless, and modest. Less experienced farmers may apply pesticides and insecticides recklessly during screening because they make bad decisions. Because of this, there have been catastrophic losses in terms of money (Kartikeyan & Shrivastava, 2021). A key component in addressing these challenges is the use of automated image processing methods for the detection of plant leaf diseases. Timely perception is essential for the effective monitoring and interdiction of plant leaf diseases and choices of agricultural products. Using the PlantDoc dataset and the GIN model, this paper aims to establish a viable approach for predicting leaf species and diseases in thirteen different types of plants. An extensive library of plant leaf photos annotated with disease and species names is available in the PlantDoc dataset. A multi-tasking object detection model, GIN is state-of-the-art. This research concentrates on utilizing Graph Isomorphic Networks (GIN) for the early detection of plant diseases, tackling significant agricultural issues. This study focuses on scalable methods for the efficient identification of various plant diseases, leveraging recent breakthroughs in machine learning. The purpose of this study is to assess how well the GIN model can identify plant diseases and species using leaf photos. The study's findings will shed light on the feasibility of disease and species prediction in leaves, as well as its potential uses in forestry and agriculture.

Literature Survey

Recent advances in machine learning and image processing have enabled autonomous agricultural disease detection. Jadhav *et al.* (2021) suggested identifying plant diseases with a Convolutional Neural Network (CNN). This method detected soybean plant diseases using pre-trained CNN models. GoogleNet and AlexNet were used for transfer learning. Despite improved results, the model's classification variety was weak. Huang *et al.* (2019) developed the Efficient Net model to categorize input into multiple labels using a CNN. CNN had hidden layers before. Plant diseases are now better identified. The model yielded poor results when evaluated using reference datasets and strong, efficient, loss-free CNN. Panchal *et al.* (2023) suggested a CNN-based Deep Learning (DL) model for accurate plant disease categorization. First, preprocessing, and then segmentation. An Artificial Neural Network (ANN) classifies things. The model recognized 93.6% of classes, but it misclassified several later on. Insufficient data also hampered the model. A hybrid CNN by Lakshmi Narayanan *et al.* (2022) improved banana plant disease classification accuracy. Kiani and Mamedov (2017) automated the detection and categorization of plant diseases using a Genetic Algorithm (GA) as the image segmentation method. The SVM classifier attained an accuracy rate of 86.55%, and the Minimum Distance Criterion with k-mean clustering had a rate of 95.72% when it came to disease classification. Maximum Distance Criterion classifier integration with evolutionary algorithm improves accuracy to 93.64%. Benzothiadiazole (BTH) prevented powdery mildew infection in wheat by interfering with many stages of the pathogen's life cycle (Zhou *et al.*, 2013). To make sure it lasts, we use a machine learning method called support vector machine (SVM). This section mostly focuses on wheat plants and strategies for disease prevention. In Wang *et al.* (2012), image processing algorithms and an

(ANN) are used to identify plant diseases early and accurately. ANNs with capacity lists for order execution performed better in trials, with 91.9% accuracy. The assessment reveals several key characteristics that require a basic method of disease detection in plants to improve the agriculture industry (Gavhale & Gawande, 2014). Plant disease detection methods include Stochastic Gradient Descent with Momentum (SGDM), K-implies bunching, SVM, and Back Propagation Neural Network (BPNN). SVM classifier for plant disease detection (Kaur & Kang, 2015). In a single unhealthy image, the background and black pixels are divided at the start with a value of 5.55. This ground-breaking investigation will identify plant pollution. This study evaluates early plant rust detection strategies (Naikwadi & Amoda, 2013). Large c-insulins Wheat leaf implanted highlights are retrieved via clustering, infection detection, sort recognition, and ID computing, (ANNs) help us do this. Similarly, (Ferentinos, 2018) designed a CNN model for illness detection in plants; they found 57 distinct groups of easy-to-understand plant-problem combinations. The findings for these classes were 99.55%, which led to the suggestion of using them for early plant disease diagnosis in real time. This study's proposed method automatically predicted treatment response for diseased plants by using GIN, a deep learning algorithm, to learn discriminative characteristics from functional connectivity. This study also attempted to find the most discriminative sick regions for treatment response prediction, which could be a predictive imaging biomarker for early treatment efficacy identification in plants.

Materials

Proposed Methods

Plants are vital for human energy generation and have nutritional and therapeutic benefits. Plant diseases can harm crop yield and economic value at any point during the farming process. In the farming industry, identifying leaf disease is vital. However, it requires significant labor, preparatory time, and extensive plant pathogen expertise. Researchers have created and tested several Machine Learning (ML) and Deep Learning (DL) algorithms for detecting plant diseases, yielding considerable results in both. This article examines the performance of GNN, GNN-LSTM, BiGRU, ELM, DNN, BiGRU-Att, LSTM-DNN, and GIN for detecting plant diseases.

Digital signal processing is an approach for obtaining fast and precise results about plant leaf diseases. It will reduce numerous agricultural issues while increasing productivity by detecting the relevant diseases. Figure (2) depicts the conceptual structure of the plant disease detection pipeline, emphasizing essential components such as preprocessing, feature extraction, and disease classification via GIN. This framework displays the orderly progression from raw picture input to disease prediction. For disease detection, an image of an infected leaf should be examined using a set of methods. Figure (2) indicates that the input image should be preprocessed before its features are retrieved based on the dataset (Dagwale & Adakane, 2023). Following that, certain classifier techniques should be applied to categorize diseases based on the specific data set.

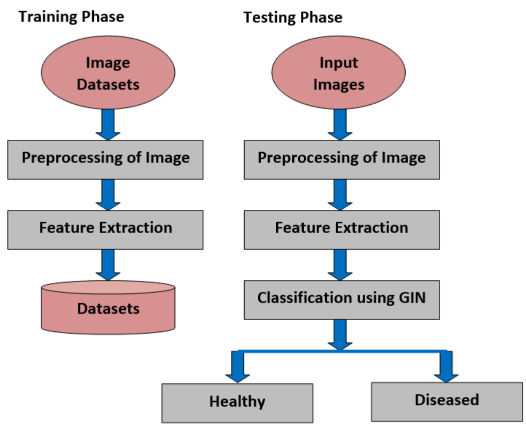

Fig. 2: Proposed model architecture

Table 1: Summary of PlantDoc dataset

Crop	Classification	Images
Blue Berry	Healthy	117
Graph	Healthy	69
	Black Rot	64
	Healthy	91
Apple	Scab	93
	Rust	89
Cherry	Healthy	57
	Leaf Blight	192
Corn	Grey Leaf Spot	68
	Rust	116
Strawberry	Healthy	96
Bell Pepper	Healthy	61
	Leaf Spot	71
Peach	Healthy	112
Squash	Powdery Mildew	130
Potato	Early Blight	117
	Late Blight	105
Soybean	Healthy	65
Raspberry	Healthy	119
Tomato	Healthy	63
	Early Blight	88
	Late Blight	111
	Spider Mite	2
	Bacterial Spot	110
	Yellow Virus	76
	Septoria Leaf Spot	151
	Mosaic Virus	54
	Leaf Mold	91

Plant Doc Dataset

The PlantDoc dataset (Uddin, 2024) shares similar classes and illnesses with PlantVillage. It is also publically available. However, the PlantDoc dataset is substantially smaller. This study employed Images from PlantDoc datasets. Table (1) summarises the datasets used in this investigation. There are a total of 2598

DOI: 10.3844/jcssp.2025.2065.2073

photos (Leygonie *et al.*, 2024; Ahmad *et al.*, 2023a). The PlantDoc collection includes images of plant illnesses collected in the field. However, the majority of the images were obtained from online sources, making the dataset very volatile. The PlantDoc collection contains images of leaves collected in their natural habitat and retrieved via internet scrapping. The collection includes 2578 pictures from 13 plant kinds, with 30 classes indicating healthy and sick leaves (Figure 3). The dataset contains 67% of photos with abnormalities.

(a) Train subset (diseased and healthy) (b) Test subset (diseased and healthy)

Fig. 3: Example images of PlantDoc dataset for both training and testing

Images of plant illnesses collected from the field and annotated for the purpose of training models to detect crop diseases from field condition photos are housed in PlantDoc (Ahmad *et al.*, 2023b). Figure (3) shows that several of the images in the dataset featured leaves that did not appear to have been captured on plants and instead looked more like images taken in a lab and that the images were generally of low quality due to being obtained from the internet. The PlantDoc collection, while globally sourced, encompasses crops and illnesses prevalent in India, rendering it relevant to actual agricultural issues in these areas.

Preprocessing

It is possible to make more accurate prediction judgments using raw photos with high-resolution (HR) images. Improved disease prediction for plants is often possible with higher-resolution images compared to lower-resolution ones (Ojo and Zahid, 2023). The creation of automated diagnostic tools is made feasible by HR pictures, which aid farmers in spotting problems early and making informed decisions.

Object detection and picture segmentation are both improved by this. When there are a lot of variances in the input photos, it might be difficult for non-learning-based deterministic image preprocessing methods, such as high-pass filters, to maintain the same level of enhancement across a wider range of images. The methods can be enhanced by making the input adaptable. Raw photos, where a high degree of unpredictability is common, may be ideal for this.

Adaptive Histogram Equalization

AHE is a method for enhancing picture contrast that selectively modifies a small area of the image (the tile). An enhancement to histogram equalization is computed and applied for each tile in order to increase contrast (HE), acting as a contrast transform function. When there are areas of a picture that are noticeably lighter or darker than the rest of the image, this method also fails to properly adjust the contrast. For photographs of plants, this could be helpful because the contrast in different parts of the image is likely to be different. Using the AHE approach, on the other hand, boosts contrast and introduces additional noise to regions of the image that are otherwise rather stable.

Image Sharpening

An essential technique for improving the overall visual impact of photographs is sharpening, which does this by raising the contrast between the image's edges. By increasing the image's high-frequency components, a high-pass filter is used to acquire a sharpening mask before sharpening the image. Following the sharpening process, the edge's gradient will be amplified. Image noise has increased, which is definitely an issue, even when very minor imperfections are shown.

CLAHE + Sharpening (CL + SH)

Applying CLAHE first and then performing an image sharpening phase is an example of a combination method that works. This method of composite picture preparation was motivated by two main ideas. Firstly, CLAHE enhances images without noise by utilizing a clip limit; however, this also restricts CLAHE's image-improving capabilities, which leaves space for future improvements. Secondly, an unrestricted comprehensive improvement is nevertheless compromised by noise when it comes to picture sharpening. Therefore, CLAHE may be able to improve and decrease noise in the image by sequentially using these two steps. The next step is to apply sharpening to the image to make it even better without distorting it.

Segmentation and Feature Extraction of Image

Pattern recognition relies heavily on feature extraction. Features selected for identification through classification play a pivotal role in pattern recognition. By comparing the intensity of photos of sick plants, researchers were able to identify the pattern of symptoms caused by the disease. What makes up the digital image is the data included in the image, which comprises the pixel-by-pixel values of color intensity. The three primary colors, red, green, and blue, make up the apparent value of a color. As a result, specific hues represent a vector system with orthogonal axes in accordance with the color space's established norms. In order to determine the efficacy of intensity-based statistical features for plant disease identification, it employed a decision tree.

The feature extraction method presented in Algorithm 1 is one such procedure. First, the algorithm creates feature vectors, which are subsequently fed into the decision tree for classification (Sabrol & Kumar, 2016). Prior to proceeding, the images are adjusted to a standard of 256 × 256. The next thing to do was alter the hue of standard images using Otsu's segmentation. Make the switch to the CIE *W Z* and *Y* color space after that. Using the images of healthy and sick plants then computed a total of ten statistical characteristics. It includes three means, three standard deviations, and three skewnesses when it comes to the exacted components *W*, Z, and *Y*, in addition to related features of the W and Z components. The final feature vector is composed of:

$$[CIE\_WZYColFea] = \left\{ \begin{array}{c} mean\left(w\right), mean\left(z\right), mean\left(y\right), std\left(w\right), std\left(h\right), std\left(a\right), \\ skwness\left(q\right), skwness\left(h\right), skwness\left(a\right), corr\left(w,z\right) \end{array} \right\}$$

Here are the color descriptions:

$$\sigma_j = \frac{1}{M}\sum_{i=1}^{M} e_{ji} \tag{1}$$

$$\mu_j = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(e_{ji}-\sigma_j\right)^2} \tag{2}$$

$$T_j = \sqrt[3]{\frac{1}{M}\sum_{i=1}^{M} e_{ji}-{\sigma_j}^3} \tag{3}$$

Where, M is the image's pixel count, and e_{ji} is the value of the $j-th$ color component of pixel i.

Here is the description of the color correlation coefficient:

$$q = \frac{\sum_w \sum_z \left(W_{wz}-\overline{W}\right)\left(Z_{wz}-\overline{Z}\right)}{\sqrt{\left(\sum_w \sum_z \left(W_{wz}-\overline{W}\right)^2\right)\left(\sum_w \sum_z \left(Z_{wz}-\overline{Z}\right)^2\right)}} \tag{4}$$

Where, $\overline{W} = mean2\left(W\right)$, and $\overline{Z} = mean2\left(Z\right)$. The color feature space offers a variety of correlation vectors, which are represented by the W and Z. Before sending the features to final classification, they are normalized. Images of diseased or uninfected plants can undergo processing normalization, which alters the intensity value range of individual pixels and applies the change to feature vectors (Sabrol & Kumar, 2016). The goal was to improve the categorization capabilities of the newly created feature vectors by focusing on their individual components. Applying Zero-Mean and Unit-Variance Normalization (MV) to an n-dimensional feature vector w obtained from the provided images of diseased or non-infected plants allowed us to normalize the features.

The normalized feature vector W is produced by MV by scaling all the W, Z, and Y components $w_j\left(j = 1, 2, \ldots\ldots.m\right)$ of w in the $CIEWYZ$colour space using the following expression:

$$W = \frac{W_{j-\sigma}}{\mu}_j, j = 1, 2, \ldots\ldots.m \tag{5}$$

Where $W = [CIE\_WZYColFea]$, μ and σ denote the mean value of the feature vector and standard deviation, respectively. By applying the MV approach, the feature vector w is converted into a random variable with a mean of zero and a variance of one. At last, we generated the vector of normalized features.

GIN Model Training

In recent times, the graph convolutional network has arisen as an attractive and potent paradigm for handling data from non-Euclidean graphs. Brain regions can be thought of as nodes in a network and connections between them as edges; this representation is a perfect fit for the human brain. The $0-1$ binarized functional connectivities (FCs) served as the edges connecting nodes, and the connections between other brain regions were the present node's characteristic (Liu & Wang, 2021). $H = (U, F)$with a feature vector Wu per node can thus be used to depict the human brain for any node $u \in U$. Within this study, H stands for the subject's created brain network, U for brain areas, and F for the correlation between them. Our objective is to create a representation vector g_H that can anticipate the label of graph H given a collection of graphs $\{H_1, \ldots.H_M\} \in H$ and their labels $\{Z_1, \ldots..Z_M\} \in Z$.

The robust GNN models learn the graph H's representation or the node feature Wu from the topological structure and the node feature. Most GNNs employ the neighborhood aggregation method (Duan *et al.*, 2023). By combining the representations of nearby nodes, this method iteratively updates the node representation. With $l-$iterations, one can learn the structure of the $l-$hop neighbors of a node. The following is one representation of the $l-$th layer of a GNN:

$$b_u^{(l)} = AGGREGATE^{(l)}\left(\left\{g_v^{(l-1)}; v \in M\left(U\right)\right\}\right) \tag{6}$$

$$g_u^{(l)} = COMBINE^{(l)}\left(g_u^{(l-1)}, b_u^{(l)}\right) \tag{7}$$

For node u on the lth layer, the feature vector is represented as $g^{(ol)}$d. $M\left(u\right)$ represents the neighbor node set with u, and we started with $g_u^{(0)} = WU$. The readout function culminates in the following transformation of the graph's node characteristics into graph features:

$$g_H = READOUT\left(\left\{g_u^{(l)}\right\}_{u \in U}\right) \tag{8}$$

Brain regions with similar topologies in a network are likely to share functional characteristics. Therefore, correctly identifying brain regions with similar architecture is crucial for investigating brain disorders. However, GNNs reach their limit when it comes to distinguishing graph structures, and the WL test method finds isomorphism between two graphs depending on the number of nodes with edges and the connectedness of edges in the two graphs. Since it is not feasible to map two separate neighborhoods to the same representation, an injective aggregation pattern is necessary for a GNN to have the same degree of power as the WL test procedure. Because it gathers neighbor nodes through the action of multiset injector functions, GIN improves the performance of graph convolutional neural networks on

DOI: 10.3844/jcssp.2025.2065.2073

homomorphic graphs. Specifically, it iteratively gathers and updates node characteristics using the following formula:

$$g_u^{(l)} = MLP^{(l)}\left(\left(1+\in^{(l)}\right)\times g_u^{(l-1)} + \sum\nolimits_{v\in M(U)} g_v^{(l-1)}\right) \tag{9}$$

The value of $\in$ can be either fixed or learnable. The human brain is like a graph, with various regions serving as nodes. In order to get the graph-level representation for the treatment response prediction, we used the following readout function:

$$g_H = \left(SUM\left(\left\{g_u^{(l)}\,|u\in U\right\}\right)|l = 0,1,2,\ldots\ldots L\right) \tag{10}$$

In which g_H stood for the graph characteristic of each subject. The last step was to use linear layers and the $\log - softmax$ algorithm to classify the features. Although crop-specific datasets are optimal for targeted applications, the GIN model's capacity to generalize across many crops and diseases illustrates its potential as a scalable solution for areas with varying agricultural practices.

Results and Discussion

The importance of conducting relevant research to sustainable agricultural development is highlighted by the increasing usage of artificial intelligence in plant disease diagnosis and other advancements in agricultural technology. Manually interprweting the symptoms of leaf diseases, such as early blight and late blight, is a laborious and time-consuming process that has a significant impact on potato yield and quality. Automated and effective diagnosis of these diseases during the budding phase can help improve potato crop output, even though it needs a high level of skill. A number of models for identifying plant diseases have been put forward in the past. This study introduces a technique that extracts useful characteristics from a dataset by fine-tuning (transfer learning) pre-trained models such as GIN.

The data was divided into an 80-20% ratio for training and testing purposes. The models Area Under the Curve (AUC), Classification Accuracy (CA), Precision (P), Recall (R), and F1-Score are displayed in Table (2).

Table 2: Comparison of Models

Models	Precision	F1 Score	AUC	CA
GNN	93.8	93.8	98.7	93.4
CNN-LSTM	90.9	90.9	98.4	90.9
BiGRU	92.6	92.7	98.6	92.7
ELM	91.5	91.5	98.8	91.5
DNN	93.6	93.7	98.3	93.8
BiGRU-Att	94.1	94.1	98.7	94.2
LSTM-DNN	90.6	90.6	98.2	90.6
GIN	95.6	95.6	98.9	95.6

The GIN model outperformed all other models in plant disease identification with a top-1 error rate of only 0.48% and a classification accuracy of 95.63%. This impressive precision demonstrates the model's strong capacity to correctly detect plant diseases. Looking at how the GIN model fared on the testing dataset during training compared to other models is shown in Figure (4). The findings confirm that the GIN model is the best option for jobs involving plant disease identification due to its efficiency and accuracy. More accurate and automated disease identification is now possible thanks to this huge leap forward in applying cutting-edge machine-learning techniques to agricultural applications.

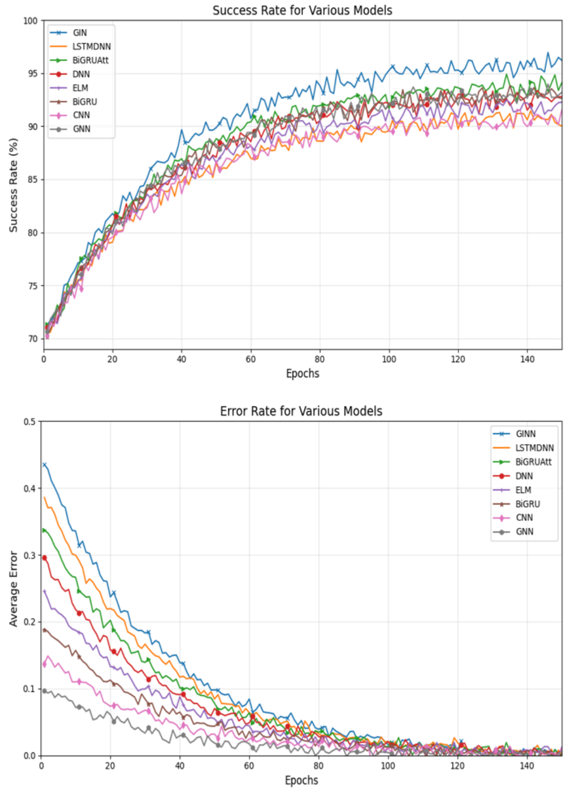

Fig. 4: Accuracy and loss values at different levels

Examining the potential for CL + SH to enhance the performance of different deep learning classifiers (GNN, CNN-LSTM, BiGRU, ELM, DNN, BiGRU-ATT, LSTM-DNN, GIN) when combined with a GIN-based method is the goal here. Consequently, using the data set in Table (3), compare the results of 8 weighted deep learning classifiers trained on both raw pictures and CL + SH, two forms of preprocessed input. With its ability to improve the functioning of deep learning classifiers, CL + SH substantially beats raw pictures in terms of ACA, as shown in the Table.

The mean F1 score over a 30-epoch training period was analyzed across all experiments, with experimental configuration parameters aggregated together. The intensity of a certain class at any given point indicates the level of uncertainty across experiments using that specific setup. Figure (5) depicts a comparison of the progression of the mean F1 score across all experiments organized by the deep learning architectures utilized. This analysis sheds light on how different model configurations affect both performance and stability during training. The full study shows how some configurations regularly reduce uncertainty, resulting in higher F1 ratings. By visualizing these trends, the analysis provides a thorough knowledge of the interaction between experimental parameters, uncertainty levels, and overall model performance across a variety of configurations.

Table 3: Different Preprocessing Techniques for GIN Classification

Preprocessing Technique	Indices	Healthy	Diseased	Overall Performance
AHE	Precision	87.2	83.6	85.4
	Recall	86.1	84.3	85.2
	F1-Score	86.7	83.9	85.3
CL+SH	Precision	95.8	95.4	95.6
	Recall	95.8	95.4	95.6
	F1-Score	95.8	95.4	95.6
SH	Precision	95.3	94.1	94.70
	Recall	92.2	96.4	94.3
	F1-Score	93.5	95.3	94.4

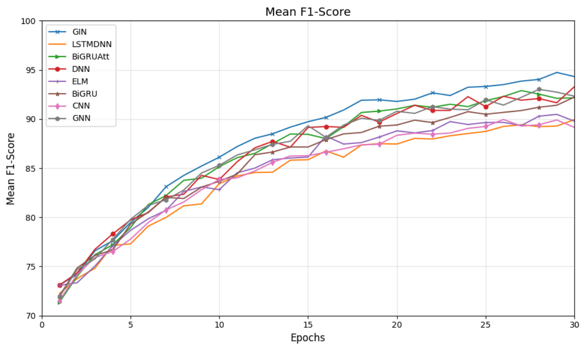

Fig. 5: Proposed model F1 score comparison

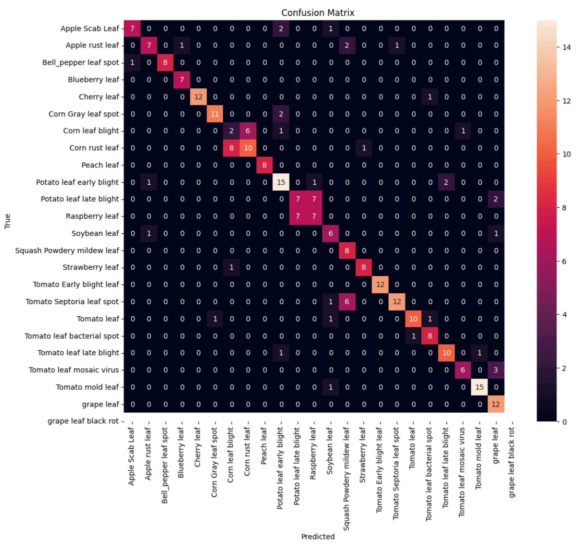

Fig. 6: Confusion Matrix for the Proposed Model (PlantDoc)

Figure (6) reveals that the GIN model outperformed others due to its effective compound scaling strategy. This strategy consistently grows network breadth, depth, and resolution, allowing to maintain of an optimal balance between model size and accuracy, exceeding standard scaling approaches used in GNN, CNN-LSTM, BiGRU, ELM, DNN, BiGRU-ATT, and LSTM-DNN architectures. Despite its deep architecture, BiLSTM had the lowest performance of the models. The smaller and less complicated PlantDoc dataset may not have completely used DNN's vast capacity, potentially resulting in overfitting. Although dense connections increase feature reuse and alleviate the vanishing gradient problem, they may have introduced duplication in feature maps for this task, resulting in lower performance relative to CNN-BiLSTM.

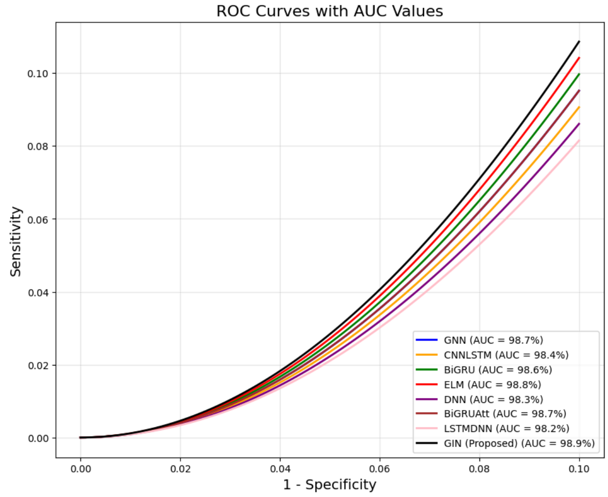

Fig. 7: ROC curve of the models

Fig. (7) shows a Receiver Operating Characteristic (ROC) curve that compares the performance of several models depending on their sensitivity (true positive rate) versus 1-specificity (false positive rate). Each curve represents a distinct model, and their Area Under the Curve (AUC) values are shown in the legend to reflect their performance. The proposed model, GIN (black curve), has the highest AUC of 98.9%, beating out other models such as ELM (98.8%), GNN and BiGRUAtt (both 98.7%), and CNNLSTM (98.4%). The ROC curves are tightly grouped, indicating competitive model performance. The legend, positioned in the bottom-right corner, is color-coded for easy identification. This visualization demonstrates the proposed GIN model's enhanced discriminative power, as evaluated by the AUC metric. The chart is well-labeled with axes ("Sensitivity" and "1 - Specificity") and a title ("ROC Curves with AUC Values") for clarity.

Figure (8) is a bar chart named "Performance Comparison," which shows the evaluation of multiple models using four metrics: CA (Classification Accuracy), Recall, F1 Score, and Precision. The models being compared are GIN, LSTM-DNN, BiGRU-Att, DNN, ELM, BiGRU, CNN-LSTM, and GNN, with each represented by a different color-coded bar. The performance scores for each metric vary from 88 to 96, indicating that the models are highly accurate and consistent. The GIN model looks to perform the best across all criteria, followed by BiGRU-Att and CNN-LSTM, both of which have competitive scores. GNN and LSTM-DNN perform relatively well, with ELM and BiGRU scoring somewhat lower than the rest. The graphic effectively depicts the relative strengths of each model in handling classification tasks, with GIN emerging as the most robust performer.

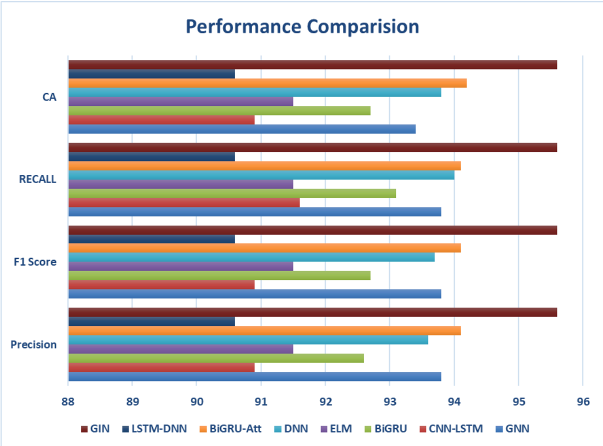

Fig. 8: Performance comparison

Conclusion

Crop diseases have recently seen a meteoric rise, thanks to both altered weather patterns and a general lack of crop immunity. As a result, farmers lose money due to the widespread destruction of crops and the subsequent decline in cultivation. Identifying and treating diseases has become a significant difficulty due to the rapid rise of both the type of diseases and the amount of knowledge that farmers have. There are telltale signs of disease in the leaves, such as similarities in texture and appearance. Therefore, the solution to this problem can be found by utilizing computer vision in conjunction with deep learning. This research presents a deep learning model that can distinguish between crop leaves that are healthy and those that are sick, using a publicly available dataset for training purposes. The model accomplishes its goal by sorting leaf pictures into a sick category according to the defect pattern. Capturing and preparing images make up the first of five stages in the suggested algorithm. Using a set-size resizer helps standardize and resize images. The preprocessing and segmentation using Otsu's approach are also part of this step. It then applied color space conversions using the segmented color images in the second phase. Eleven color descriptors for five hues are calculated in the third step. We next ran the features that were retrieved through eight separate classifiers. Finally, the accuracy of recognition was assessed.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions to update the manuscript.

Funding Information

This research received no external funding.

Author's Contributions

All authors equally contributed to this study.

Ethics

This manuscript is an original work. The authors declare that there are no ethical concerns associated with this submission

Conflict of Interest

The authors have no competing interests to declare relevant to this article's content.

References

Ahmad, A., Gamal, A. E., & Saraswat, D. (2023a). Toward Generalization of Deep Learning-Based Plant Disease Identification Under Controlled and Field Conditions. *IEEE Access*, *11*, 9042–9057. https://doi.org/10.1109/access.2023.3240100

Ahmad, A., Saraswat, D., & El Gamal, A. (2023b). A Survey on Using Deep Learning Techniques for Plant Disease Diagnosis and Recommendations for Development of Appropriate Tools. *Smart Agricultural Technology*, *3*, 100083. https://doi.org/10.1016/j.atech.2022.100083

Dagwale, S. S., & Adakane, P. (2023). Prediction Of Leaf Species & Disease Using Ai For Various Plants. *International Journal for Multidisciplinary Research*, *5*(3), 1–5.

Duan, J., Li, Y., Zhang, X., Dong, S., Zhao, P., Liu, J., Zheng, J., Zhu, R., Kong, Y., & Wang, F. (2023). Predicting Treatment Response in Adolescents and young Adults with Major Depressive Episodes from fMRI Using Graph Isomorphism Network. *NeuroImage: Clinical*, *40*, 103534. https://doi.org/10.1016/j.nicl.2023.103534

Ferentinos, K. P. (2018). Deep Learning Models for Plant Disease Detection and Diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Gavhale, Ms. K. R., & Gawande, Prof. U. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering*, *16*(1), 10–16. https://doi.org/10.9790/0661-16151016

Huang, S., Liu, W., Qi, F., & Yang, K. (2019). Development and Validation of a Deep Learning Algorithm for the Recognition of Plant Disease. *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Zhangjiajie, China. https://doi.org/10.1109/hpcc/smartcity/dss.2019.00269

DOI: 10.3844/jcssp.2025.2065.2073

Jadhav, S. B., Udupi, V. R., & Patil, S. B. (2021). Identification of Plant Diseases Using Convolutional Neural Networks. *International Journal of Information Technology*, *13*(6), 2461–2470. https://doi.org/10.1007/s41870-020-00437-5

Kartikeyan, P., & Shrivastava, G. (2021). Review on Emerging Trends in Detection of Plant Diseases using Image Processing with Machine Learning. *International Journal of Computer Applications*, *174*(11), 39–48. https://doi.org/10.5120/ijca2021920990

Kaur, R., & Kang, S. S. (2015). An Enhancement in Classifier Support Vector Machine to Improve Plant Disease Detection. *2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)*, 190–194. https://doi.org/10.1109/mite.2015.7375303

Khalid, M. M., & Karan, O. (2023). Deep Learning for Plant Disease Detection. *International Journal of Mathematics, Statistics, and Computer Science*, *2*, 75–84. https://doi.org/10.59543/ijmscs.v2i.8343

Kiani, E., & Mamedov, T. (2017). Identification of Plant Disease Infection Using Soft-Computing: Application to Modern Botany. *Procedia Computer Science*, *120*, 893–900. https://doi.org/10.1016/j.procs.2017.11.323

Lakshmi Narayanan, K., Santhana Krishnan, R., Harold Robinson, Y., Golden Julie, E., Vimal, S., Saravanan, V., & Kaliappan, M. (2022). Banana Plant Disease Classification Using Hybrid Convolutional Neural Network. *Comput. Intell. Neurosci*, *2022*, 1–11. https://doi.org/10.1155/2022/9153699

Leygonie, R., Lobry, S., & Wendling, L. (2024). Can we Detect Plant Diseases without Prior Knowledge of their Existence? *International Journal of Applied Earth Observation and Geoinformation*, *134*, 104192. https://doi.org/10.1016/j.jag.2024.104192

Liu, J., & Wang, H. (2021). Graph Isomorphism Network for Speech Emotion Recognition. *Interspeech 2021*, 3405–3409. https://doi.org/10.21437/interspeech.2021-1154

Naikwadi, S., & Amoda, N. (2013). Advances in Image Processing for Detection of Plant Diseases. *International Journal of Application or Innovation in Engineering & Management*, *2*(11), 329-342.

Ojo, M. O., & Zahid, A. (2023). Improving Deep Learning Classifiers Performance via Preprocessing and Class Imbalance Approaches in a Plant Disease Detection Pipeline. *Agronomy*, *13*(3), 887. https://doi.org/10.3390/agronomy13030887

Panchal, A. V., Patel, S. C., Bagyalakshmi, K., Kumar, P., Khan, I. R., & Soni, M. (2023). Image-based Plant Diseases Detection using Deep Learning. *Materials Today: Proceedings*, *80*, 3500–3506. https://doi.org/10.1016/j.matpr.2021.07.281

Sabrol, H., & Kumar, S. (2016). Intensity Based Feature Extraction for Tomato Plant Disease Recognition by Classification Using Decision Tree. *International Journal of Computer Science and Information Security (IJCSIS)*, *14*(9), 622–626.

Sharma, A. (2024). Plant Disease Detection Using Image Processing and Machine Learning. *CCIS*, 69–85. https://doi.org/10.1007/978-3-031-58953-9_6

Sharma, D., Kapoor, N., & Sood, D. (2023). Plant Disease Detection Techniques: A Survey. *Precision Agriculture for Sustainability*, 13. https://doi.org/10.1201/9781003435228-19

Sunil, C. K., Jaidhar, C. D., & Nagamma, P. (2022). Cardamom Plant Disease Detection Approach Using EfficientNetV2. In *IEEE Access* (Vol. 10, pp. 789–804). https://doi.org/10.1109/access.2021.3138920

Uddin, A. H. (2024). Plant Doc Dataset. Kaggle. *Kaggle*. https://www.kaggle.com/datasets/abdulhasibuddin/plant-doc-dataset

Wang, H., Li, G., Ma, Z., & Li, X. (2012). Image Recognition of Plant Diseases Based on BackPropagation Networks. *2012 5th International Congress on Image and Signal Processing*, 894–897. https://doi.org/10.1109/cisp.2012.6469998

Zhou, R., Kaneko, S., Tanaka, F., Kayamori, M., & Shimizu, M. (2013). Early Detection and Continuous Quantization of Plant Disease Using Template Matching and Support Vector Machine Algorithms. *2013 First International Symposium on Computing and Networking*, 389–393. https://doi.org/10.1109/candar.2013.52